Maryrose B. Caulfield-Sloan
Mary F. Ruzicka
Planning and Changing
Vol. 36, No. 3&4, 2005, pp. 157–175

THE EFFECT OF TEACHERS' STAFF DEVELOPMENT IN THE USE OF HIGHER-ORDER QUESTIONING STRATEGIES ON THIRD GRADE STUDENTS' RUBRIC SCIENCE ASSESSMENT PERFORMANCE

Introduction

Research suggests that aspects of teaching effectiveness make the difference in how students perform. Successful teachers tend to be those who employ a range of teaching strategies and interactive styles to meet the needs of their learners. These effective teachers utilize different instructional goals, topics, and methods (Doyle, 1985). Research further demonstrates that teachers' abilities to structure material, ask higher-order questions, use student ideas, and probe student comments have also been found to be important variables in what students learn (Darling-Hammond, Wise, & Pease, 1983; Good & Brophy, 1986; Rosenshine & Furst, 1973).

A current and urgent problem is how to train primary grade teachers in science. These teachers possess only a general science knowledge, which may have been acquired through a methods course, but are not specifically trained in process thinking or the use of the scientific method. This is evidenced by numerous teacher self-reports, as well as observation by the senior author who was the district elementary school science specialist.

One specific strategy that elementary school science teachers need to learn is effective questioning. The primary author has often witnessed teachers demonstrating a difficulty relinquishing control of the learning process over to students. Thus, teachers need to be guided toward practices by which students can own the thinking process rather than merely being recipients of information. This makes the teacher the "guide on the side, not the sage on the stage."

Due to the pressure of time constraints in the classroom, educators need to move science education from rote memory to active thinking. It is important to incorporate a strategy that teachers can easily use and that will not lend itself to personal interpretation on their part. This method needs to be modeled to teachers via professional development and be reproducible across a range of classrooms in a district.

Staff development takes on a greater significance in light of these needs. As Dennis Sparks (1997) describes it, "For too many teachers, staff development is a demeaning, mind-numbing experience in which they passively 'sit and get.' As one observer put it, 'I hope I die during an in-service session, because the transition between life and death would be so subtle'" (p. 21).

The type of staff development necessary to improve student achievement is not the type of in-service where elementary teachers just attend a workshop to learn a specific activity to be used when teaching a particular concept. Rather, a comprehensive instructional strategy is called for, one designed to enhance student comprehension and mastery

for increased student performance. This method of in-service is one in which staff members are trained in the use of a precision instrument, such as higher-order questioning, that heightens the significance and expands the learning potential of all activities and concepts, not just one particular topic. The use of a higher-order questioning process, designed from Bloom's taxonomy, formalizes the connection among the specific questions asked by the teacher. This process guides practitioners and helps them to assess the comprehension and mastery of the students, which lead directly to their performance outcomes.

Benjamin Bloom (1956) created a taxonomy for categorizing the level of abstraction of questions that commonly occur in educational settings. This taxonomy sets up a useful series of steps that identify increasing degrees of abstraction. Bloom's taxonomy is shown in Table 1. Under each category in the table are a series of verbs that can be utilized in questions related specifically to that level of abstraction. Training elementary teachers to use this taxonomy enables them to identify which type of questions lead to higher-level thinking and responses in students.

> Upon examination of Benjamin Bloom's cognitive domain contained in his Taxonomy of Educational Objectives, teachers are reminded that the classification levels of the cognitive domain, namely knowledge, comprehension, and application, are skills of recall and recognition, whereas analysis, synthesis, and evaluation comprise higher energy intellectual skills…. Because teachers' questions are used to solicit learner participation, their questions should serve as quality demonstrations that lead to the enhancement of students' ability to self-interrogate at all levels of Bloom's taxonomy. (Williamson, 1998, p. 31)

Table 1

Bloom's Taxonomy

Convergent			Divergent		
Lower-order thinking			Higher-order thinking		
Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Tell, list	Translate	Use	Examine	Create, combine	Judge, decide
choose	reword	solve	dissect	build, compile	rate, prioritize
arrange	expand	apply	divide	make, structure	appraise, assay
name	transform	employ	take apart	reorder, blend	rank, weigh
locate	retell, restate	utilize	investigate	reorganize	accept, reject
repeat	infer, define	make	discuss	cause, develop	determine
quote	explain	use of	uncover	produce	assess, referee
point to	outline	mobilize	simplify	compose, yield	umpire, arbitrate
check	annotate	manipulate	deduce	construct, effect	rule, award
recite	project	practice	conclude	generate, evolve	criticize
underline	propose		extract	form, constitute	censure, settle
identify	calculate			originate	classify, grade

Note. Adapted by Caulfield-Sloan (2001, p. 81) from Bloom (1956).

Purpose

The purpose of this study was (a) to instruct teachers in the use of these levels of questions, (b) to have the teachers implement the questions in their classrooms, and (c) to assess the level of performance among students educated with this style of teaching.

Subjects

The subjects of this study were 120 third grade students randomly chosen from a total population of 600 students. Designation of experimental and control subjects was based on the selection of these students' teachers (27 total) to be trained in the methodology of the study. Teachers were placed into either an experimental or control group by being matched on their range of teacher background and experience. From each of these two teacher groups, sixty students responses were then randomly chosen. These student groups were matched on the basis of I.Q., academic performance, and socio-economic background.

Procedure

Overview

A workshop was designed on the use of effective, higher-order questioning strategies in science for third grade teachers across the district. These workshops were conducted in the district by the elementary science staff developer (Caulfield-Sloan, 2001). The credibility and expertise of this person had already been established with the third grade teachers. This was a familiar role in the district for the primary researcher and was an extension of the trust relationship already in place as the research process occurred. This researcher occupied a line position with teachers in this district and already had the ability to enter teachers' classrooms to observe and aid in the teaching process since this activity was part of her job description. She was in no way responsible for supervising or evaluating teachers, and there was no existing feeling of coercion within her presence in the learning environment.

As a routine part of the staff development process within the district, 14 third grade teachers were given the treatment (workshop) at one time, and 13 third grade teachers were given the treatment at a later time due to the availability of substitute teachers to cover classes. Eventually, 27 third grade teachers received the training.

Teachers were fully informed of all aspects of the study. After agreeing to participate, the third grade teachers were asked to complete a survey on their backgrounds, including academic and teaching experiences and perceptions. The results of these surveys, which are included in Table 2, along with the results from pre-workshop classroom observations, were used to determine the experimental and control groups. Teachers were matched on the basis of this information.

Table 2

Results of Teacher Questionnaire on Educational and Teaching Background

Teacher perception of teaching style	Teacher perception of science teaching style	Experimental or control group	No. of years teaching	No. of years teaching third grade	No. of teaching certificates	No. of post-graduate credits	No. of post-graduate degrees	No. of science courses taken
Notes oral multi-sensory	Brain-storming oral written	Experimental	27	14	1	9	0	0
Multiple intelligence Multi-sensory	Talk Experiment Write	Control	19	7	2	30	0	0
Multi-sensory creative	Discussion hand-on	Experimental	27	11	1	48	1	1
Holistic Socratic	Include in other subjects	Experimental	25	17	1	84	1	2
Structured Flexible	With reservation	Experimental	9	1	1	18	0	2
Variety of inst. strat. for diverse learners	Hands-on diagram literature	Experimental	24	20	1	37	0	3
Multi-sensory	No response	Control	26	23	1	48	0	1
Multi-sensory	Hands-on demos charts books movies	Experimental	17	7	1	48	0	3
Traditional	Traditional hands-on	Control	31	30	1	48	0	2
Traditional manipulatives, Models, graphs, charts	No response	Control	20	5	1	0	0	0
Hands-on exploration open-ended	Hands-on lecture note experiment	Control	5	3	1	0	0	0

(continued)

Table 2 *(continued)*

Teacher perception of teaching style	Teacher perception of science teaching style	Experimental or control group	No. of years teaching	No. of years teaching third grade	No. of teaching certificates	No. of post-graduate credits	No. of post-graduate degrees	No. of science courses taken
Caring flexible	Experiment discussion	Control	24	7	1	0	0	0
Student-centered organized basic skills	Hands-on	Experimental	23	19	1	82	1	2
Enthusiastic	Hands-on experiment	Experimental	< 3	< 3	1	0	0	2
Teacher lead discussion Videos, hands-on	Videos, dittos, hands-on	Control	7	< 1	1	48	0	4
No particular style	Lessons video notes hands-on	Control	< 6	3	2	48	1	2
No response	No response	Experimental	23	15	3	48	1	1
Traditional hands-on, higher level thinking	Not enough time to teach science	Control	5	5	1	0	0	0
No response	Inquiry Method	Control	< 1	< 1	1	0	0	0
No response	Hands-on	Experimental	19	19	1	32	0	0
Relaxed with discussions	Laser disc worksheets open-ended	Control	25	9	1	0	0	2
No response	Laser disc	Experimental	14	7	4	48	1	0
Traditional	Not sure new to grade 3	Control	6	< 1	3	48	1	0

Observations conducted of all of the teachers prior to the beginning of the research process determined that they had not been using higher-order questioning strategies in their classrooms prior to the workshop. Results of these observations are included in Table 3.

Table 3

Results of Pre-Workshop Teacher Observation

Teacher question low or high	Pupil response	Patterns of teacher question	Teacher follow-up	Teacher position in classroom	Desk arrangement in room	Number of students	Teacher vs pupil participation
Low	Rote	Student hands up	None	In front of class	Paired rows	25	Teacher dominant
Low	Rote	Student hands up	None	In front of class	Paired rows	25	Teacher dominant
Low	Rote	Student hands up	None	In front of class	Paired rows	25	Teacher dominant
Low middle	Rote and some open-ended	Encouraged all pupils	Some	Through class	Teams	18	Fifty/fifty
Low	Rote	Student hands up	None	Sitting in front of class	Paired rows	18	Teacher dominant
Low	Rote	Student hands up	None	In front of class	Horse shoe	20	Teacher dominant
Low	Rote	Student hands up	None	Through class	Paired desks	20	Teacher dominant
Low	Rote	Student hands up	None	Sits at desk	Single Pairs Triplets	18	Teacher dominant
Low	Rote	Team answers	Some	Through class	Teams	20	Peer coaching
Low	Brainstorming	Student hands up	Some	Through class	Single Pairs Triplets	19	Fifty/fifty
Low	Rote	Encouraged all pupils	None	Through class	Paired rows	20	Teacher dominant
Low	Rote	Student hands up	None	In front of class	Paired rows	21	Teacher dominant
Low	Rote	Student hands up	None	In front of class	Paired rows	20	Teacher dominant

(continued)

Table 3 *(continued)*

Teacher question low or high	Pupil response	Patterns of teacher question	Teacher follow-up	Teacher position in classroom	Desk arrangement in room	Number of students	Teacher vs pupil participation
Low	Rote	Picked names	None	In front of class	Groups	23	Teacher dominant
Low	Brain-storming	Student hands up	None	Through class	Horse shoe	20	Fifty/fifty
Low	Brain-storming	Student hands up	None	In front of class	Groups	23	Teacher dominant
Low	Rote	Student hands up	None	In front of class	Groups	23	Teacher dominant
Low	Rote	Student hands up	None	In front of class	Rows	23	Teacher dominant
Low	Rote	Picked names	None	In front of class	Paired rows	22	Teacher dominant
Low	Rote	Student hands up	None	In front of class	Horse shoe	21	Teacher dominant
Low	Rote	Student hands up	None	Through class	Paired rows	23	Teacher dominant
Low	Rote	Student hands up	None	In front of class	Rows	21	Teacher dominant
Low	Rote	Student hands up	None	Through class	Groups	19	Teacher dominant
Low	Rote	Called on each other	None	In front of class	Paired rows	18	Teacher dominant
Low	Rote	Picked names	None	In front of class	Paired rows	20	Teacher dominant
Low	Rote	Student hands up	None	Through class	Paired rows	19	Teacher dominant
Low	Rote and brain-storming	Student hands up	None	In front of class	Paired rows	19	Teacher dominant

Training Instrument

To illustrate the use of these questioning strategies in a classroom setting for the workshop, a lesson in the form of a science experiment was taught to a group of third grade students and recorded on video. These students were from the third grade class of the school year just prior to the start of this research. On this video, the primary researcher modeled the desired behavior with students, utilizing a variety of questioning strategies during the experiment, and then conducted a follow-up session with the third graders to identify information learned by each child. The experi-

mental teachers observed this process on the video.

The teacher observers were given worksheets with tally areas for specific types of questions listed, reflecting levels of higher-order thinking skills (Bloom, 1956). The teacher observers recorded the types and frequencies of questions posed to the students throughout the videotaped lesson. This question asking tally sheet is shown in Figure 1.

Question Asking Tally Sheet

Teacher Code ____________________ School Code ____________________

	Convergent Thinking			Divergent Thinking		
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluate
Num- ber of times asked	Tell, list choose arrange name locate repeat quote point to check recite underline identify	Translate reword expand transform retell, restate infer, define explain outline annotate project propose calculate	Use, solve apply employ utilize make use of mobilize manipulate practice	Examine dissect divide take apart investigate discuss uncover simplify deduce conclude extract	Create, com- bine, build compile make, struc- ture, reorder blend, reor- ganize, cause develop, pro- duce, com- pose, yield construct effect, gener- ate, evolve form, consti- tute, originate	Judge, decide rate, priori- tize, appraise assay, rank weigh, accept reject, deter- mine, assess referee umpire, arbi- trate, rule award, criti- cize, censure settle, classi- fy, grade
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
. . .						
30						
Total						

Figure 1. Question asking tally sheet.

After viewing the video of the students' laboratory session, there was a debriefing session. Here, the teachers were able to discuss their observations and the implications of these observations with the researcher.

Teachers were given a quantity of these tally sheets used during the workshop process. They were instructed to practice these acquired skills in the time following the workshop.

A schedule for classroom visitations by the primary researcher was established where the teachers would conduct a post-workshop lesson themselves. This researcher visited their classes while they conducted the lesson themselves, and she logged data about the questioning strategies employed by the teachers during this actual lesson. The results of these observations are included in Table 4. Participants were reminded that this was not evaluation of them, but rather a research-based opportunity to assess the effectiveness of the in-service process.

Table 4

Question Asking Tally Sheet Frequencies for the Experimental Group of Teachers

	Convergent thinking			Divergent thinking		
Teacher	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
1.	2	4	2	5	3	7
2.	5	6	3	4	3	5
3.	3	3	2	5	6	6
4.	4	4	3	4	4	5
5.	5	4	5	6	4	5
6.	3	3	2	4	4	6
7.	4	3	4	3	0	0
8.	7	3	1	3	0	0
9.	6	4	3	4	3	4
10.	1	1	3	4	5	5
11.	5	2	1	5	3	4
12.	2	1	2	4	4	4
13.	2	3	3	2	4	3
14.	3	2	2	5	5	7

It should be noted that, for fairness to all third grade teachers and students in the district, a second workshop was then conducted with the control group of third grade teachers. The workshop was the same higher-order questioning process as the one conducted with the experimental group. The same video experiment of the researcher with the group of students was viewed and served as the model for the control group teachers. Tally sheets were given to the control group members to record types and frequencies of questions used. The researcher made follow-up visits to the

control classrooms to observe a lesson of the control teachers' choosing since the lesson performed by the experimental teachers had already been taught. Observations and recording of data on the use of higher-order questioning by teachers in the control groups during science classes was again performed. This follow-up workshop was not part of the research but was performed to provide the same staff development to all third grade teachers in the district, not just the experimental group.

Student Assessment

After the experimental group of teachers completed the process of teaching the science lesson in their classroom, and the primary researcher observed and documented it, third grade students, both from the experimental and control teachers, were assessed. All third grade students in the district, including the control groups whose teachers had not yet attended the staff development workshop, were taught the same topic in science at this point, as a function of the district's curriculum time line. This assessment was an open-ended, rubric assessment involving the use of higher-order thinking responses on the part of the third graders. The rubric scores ranged from a 0 indicating no proficiency on the topic, to 1 indicating only partial proficiency, to 2 suggesting proficiency, to 3 indicating advanced proficiency. These categories are the same as the Elementary School Proficiency Assessment administered to all third and fourth graders in the state during the month of April and May each year in New Jersey. The open-ended question and rubric are shown in Figure 2.

Third Grade Open-Ended Question	
How do the roots of a plant act like a drinking straw? How do the roots of a plant act differently than a drinking straw? Use what you have learned about plants to explain your answer.	
Rubric for Open-Ended Question	
3-point response	Student response is reasonably complete, clear, and satisfactory. Student must include three or more of the following items in his/her answer: 1. Alike: Both a straw and roots carry liquids inside them. 2. Alike: Both a straw and roots bring liquids up from a lower place to a higher place. 3. Different: Roots need to have a much narrower diameter than a straw to work. 4. Different: A straw needs suction to work and roots use capillary action to draw water up inside the root without any suction.
2-point response	Student response has minor omissions and/or some incorrect or irrelevant information. Student includes two of the four items listed above.
1-point response	Student response includes some correct information, but most information included in the response is either incorrect or not relevant. Student includes one of the four items listed above.
0-point response	Student attempts the task but the response is incorrect, irrelevant, or inappropriate.

Figure 2. Third grade open-ended question, and rubric for assessing it.

The primary researcher collected all the rubric assessment papers from the 27 third grade classes. All papers had coded front sheets to conceal any identifying information about the student answering the question, including demographic information and whether the respondent was from the experimental or control group. The researcher also tallied the data from the observation sheets on the frequency and type of higher-order questions asked by teachers from the experimental groups during the post-treatment observation sessions by the researcher.

Five student responses from each of the 13 control classes and five student responses from each of the 14 experimental classes were chosen randomly by use of a table of random numbers and then matched for I.Q., academic, and socio-economic background.

I.Q. was identified by student results on the Cognitive Abilities Test (Riverside, 1993). Students were included if they fell within the I.Q. range of standard age scores (SAS) between 85 and 115. This range was determined from the average normal range of standard age scores in the standard normal curve of the population distribution. The population mean for the standard age scores of I.Q.s is 100 with a standard deviation of 15.

Academic background was determined by eliminating students who received basic skills instruction or who were in the gifted and talented program. Basic skills instruction students were identified as those students who received a National Percentile Rank of below 25% for mathematics, reading, or language on the TerraNova Basic Multiple Assessment Plus (CTB/McGraw-Hill, 1997), the established district guideline for providing basic skills instruction to a particular student. The range of I.Q. chosen eliminated those students who were in the gifted and talented program, since a base standard age score for I.Q. of 120 is required for admission into that program in the district.

Socio-economic background was determined by the use of the Free- or Reduced-Price Lunch eligibility program. Students who receive either free- or reduced-price lunch require parents or legal guardians to provide proof of eligibility for this program which is determined by economic need.

As stated earlier, there had originally been 14 experimental teachers. This number was reduced to 12 when the primary author, for ethical reasons as staff development specialist, decided to step in to instruct the class when it was observed that these teachers were not utilizing the higher-levels of questions in the follow-up experiments. These two classes were then dropped from the study because the research protocol was discontinued. That is, because these two teachers did not use higher-order questioning strategies, their students' assessment would not be reliable or valid and thus should not be included in the analysis. The two sets of five student responses that had been randomly selected and matched to those of the other 12 experimental classes were then eliminated, resulting in an experimental sample size *N* of 60. One of the original 13 control classes was eliminated from the study when the regular classroom teacher became ill and a substitute took over within the course of the study. This allowed

for 12 control classes with five student responses from each being randomly selected and matched, for a control *N* of 60. The frequency of rubric scores for the control group and the frequency of rubric scores for the experimental group are presented in Table 5.

Table 5

Comparison of Frequencies of Rubric Scores for Control versus Experimental Groups

Control rubric results			Experimental rubric results		
Rubric score	Frequencies	Percentage	Rubric score	Frequencies	Percentage
0	24	40%	0	4	6.7%
1	30	50%	1	19	31.7%
2	6	10%	2	22	36.7%
3	0	0%	3	15	25.0%

The frequencies of low (0 and 1) results and high (2 and 3) results for each group are presented in Table 6.

Table 6

Comparison of Frequencies of Lower-Order Thinking (0-1) and Higher-Order Thinking (2-3) Rubric Scores for Control versus Experimental Groups

Control rubric results			Experimental rubric results		
Rubric score	Frequencies	Percentage	Rubric score	Frequencies	Percentage
0-1	54	90%	0-1	23	38.4%
2-3	6	10%	2-3	37	61.7%

Data Analysis

This research study employed a mixed method, quasi-experimental approach. Results of the qualitative components of the study are shown as follows:

1. A pre-workshop survey of teacher background and teaching style (see Table 2);
2. Pre-workshop observations of teacher instructional styles conducted by the researcher (see Table 3);
3. Question-asking tallies collected by the researcher during post-workshop observations of the teachers in their classrooms (see Table 4).

The quantitative component of the study involved an analysis of the open-ended responses of the students following a science lesson taught to both the experimental and control classes. These responses were scored by the use of a rubric (see Figure 2). The rubric scores were calculated and then analyzed by a chi-square analysis. By examining the data for frequency of results and calculating chi-square analysis of this information, the appearance of specific results takes on a meaning that helps to interpret and explain what was learned from this study.

The overall rubric scores of the control group were compared with the overall rubric scores of the experimental group (see Table 5). The control group had a much higher frequency of non-proficient zero responses, 24, than the experimental group, which had only four non-proficient zero responses. The control group had 30 partially-proficient one responses compared to the experimental group which had only 19 rubric scores of one. Compared to the experimental group where 22 students received a proficient rubric score of two, the control group had only six students who scored a proficient two on the rubric assessment.

Finally, only the experimental group had students who achieved a level of advanced proficiency on the rubric assessment. The experimental group had 15 students who received a rubric score of three. No control group members scored in the advanced proficient area on the rubric assessment.

A two-way chi-square test was performed to determine the significance of this difference in frequencies of each category in the experimental and control groups. A value for chi-square of 39.99 was calculated. This value exceeded the critical value of chi-square of 16.27 at three degrees of freedom (*df*) with a p value <.001, indicating that the difference in frequencies observed between the experimental and control groups was not by chance (see Table 7).

Table 7

Chi-square Analysis of Individual Rubric Scores for the Experimental versus the Control Groups of Students

	Experimental		Control		Total observed frequencies
	Observed	Expected	Observed	Expected	
Frequencies of 0 results	4	(14)	24	(14)	28
Frequencies of 1 results	19	(25)	30	(25)	49
Frequencies of 2 results	22	(14)	6	(14)	28
Frequencies of 3 results	15	(8)	0	(8)	15
Total observed frequencies	60		60		120

Note. $\chi^2 = 39.99$. $df = 3$. $p < .001$.

The data were further evaluated for the difference in frequencies between the experimental and control groups of students who scored in

the low range (non- and partially proficient 0–1) on the rubric assessment and students who scored in the high range (proficient and advanced proficient 2–3) on the rubric assessment. A difference was found between the control and experimental groups when low and high frequencies were examined (see Table 6). Only 23 experimental students received low rubric scores of zero and one indicating non- and partial proficiency on the rubric assessment. Fifty-four students in the control group received low results. In the high range of performance (proficiency and advanced proficiency), 37 experimental students received high rubric scores of either two or three while only six control students received high performance rubric scores.

A two-way chi-square test was performed to determine the significance of this difference in frequencies of low and high categories in the experimental and control groups. This time, a value for chi-square of 45.8 was calculated. This value exceeded the critical value of chi-square of 10.83 at one degree of freedom (*df*) with a *p* value <.001, indicating that the difference in frequencies observed between low performers and high performers in the experimental and control groups was not by chance (see Table 8).

Table 8

Chi-square Analysis of High and Low Rubric Scores for the Experimental versus the Control Groups of Students

	Experimental		Control		Total observed frequencies
	Observed	Expected	Observed	Expected	
Frequencies of low (0–1) results	23	(39)	54	(39)	77
Frequencies of high (2–3) results	37	(22)	6	(22)	43
Total observed frequencies	60		60		120

Note. $\chi^2 = 45.8$. $df = 1$. $p < .001$.

Limitations

1. The study was performed in the primary researcher's school system where she conducted the workshop and the assessment. The possibility for bias existed.
2. All data were collected and analyzed by one researcher who was also the workshop trainer, which is a role she was also employed to perform.
3. This study involved an N=120 students.
4. The research occurred over a period of four months, a relatively short period of time in pedagogical terms.
5. Specific demographic groups were studied together.

6. This research was limited to third grade teachers and student responses.
7. The instructional procedure, used in the form of a rubric analysis, was not standardized.
8. The time of day for the teacher training, with some teachers receiving training in the morning and some in the afternoon, could have affected the learning ability of some teachers.
9. Two teachers from the experimental group, and one from the control group, were dropped from the study.

Discussion

Qualitative Data

Pre-workshop observations of all 27 third grade teachers revealed that the predominant instructional practice being used at the time of the research was a traditional, teacher dominated, rote style of teaching. Students were passive learners. Prior to the workshop, third grade teachers were observed asking most of their questions from the lower levels of Bloom's taxonomy (1956) in the areas of knowledge, comprehension, and application. Only one staff member in 27 asked several questions from the analysis level in areas not integral to the content portion of instruction. Observations of the experimental staff member group made during the science lesson following the staff development training revealed that these teachers were asking increased numbers of higher-order questions in the areas of analysis, synthesis, and evaluation (Bloom, 1956).

Pre-workshop observations of staff revealed that the questions teachers posed from the lower levels of Bloom's taxonomy (1956) elicited only rote, convergent answers that required only content, single concept information. Following the staff development intervention, observations made during the follow-up science lessons of the experimental group of teachers revealed higher-order responses from students. Students responded with answers requiring process thinking in the analysis, synthesis, and evaluation levels of Bloom's taxonomy (1956).

Quantitative Data

The mean rubric score for the control group was .7 and the mean rubric score for the experimental group was 1.8, which represented a difference of 1.1. This indicated that the successful training of the experimental group of teachers in the use of higher-order questions produced a result of an entire rubric score gain in achievement for the experimental group of third graders. The control mean of .7 places the mean score at about one, which is in the lower end of the performance scale and indicates that the mean performance of students in the control group was only partially proficient. The mean rubric score of 1.8 for the experimental

group of third graders places them at about 2, which is proficient. This indicates that the experimental group of third graders demonstrated a higher level of thinking in their responses than did the control group.

There was a difference in the frequencies of rubric responses between these two groups. The control group of students had 40% 0 responses, 50% 1 responses, 10% 2 responses, and 0% 3 responses. The rubric responses were also compared on the basis of low 0-1 (non-proficient) responses and high 2-3 (proficient) responses. The control group had 90% low responses and only 10% high responses. (See Tables 5 and 6)

By contrast, there was a dramatic difference in the frequencies of responses for the experimental students. The experimental group of third graders had only 6.7% 0 responses, 31.7% 1 responses, 36.7% 2 responses, and 25% 3 responses. The rubric responses for the experimental group in the low range were only 38.4%, while 61.7% scored in the high range. (See Tables 5 and 6)

The frequencies of rubric scores for each category demonstrate a higher number of lower end rubric responses for the control group versus a higher number of higher end rubric responses for the experimental group. These comparative differences may reflect the specific level of questions utilized by the control teacher group versus the experimental teacher group. These comparisons further may demonstrate the level of thinking generated by the control group of third graders versus the experimental group of third graders. The control group of students responded to the open-ended question with a much higher frequency of lower-order thinking (Bloom, 1956). This reflects a more rote response to the question. The experimental group of students responded to the open-ended question with a much higher frequency of higher-order thinking (Bloom, 1956).

These differences in the frequencies of responses between the control and the experimental groups also reflect a difference in the level of mastery of the material being presented. The 0 and 1 rubric response correspond to non-proficiency and partial proficiency respectively. Table 6 shows an overwhelming 90% of control responses were within this range. The predominant adherence to rote answers to the open-ended question demonstrated the inability of these students to access the higher-levels of thinking required to respond effectively to the open-ended question.

Conversely, Table 6 also shows a dramatic 61.7% of the experimental students responded in the proficient and advanced proficient levels. These rubric scores correspond with the upper levels in Bloom's taxonomy (1956). The students who received instruction by teachers successfully trained in the use of higher-order questions were able to extend their thinking well beyond rote, convergent responses to the divergent thinking required for open-ended questions (Cardellichio & Field, 1997; Gagne, 1965).

The significance of these results was addressed through the use of the chi-square statistic as is shown in Tables 7 and 8.

Conclusions

From the pre-workshop observations and the teacher questionnaires, it was clear that the third grade teachers in this study did not possess the range of questioning skills demonstrated in Bloom's taxonomy (1956) although the teachers were an experienced group who had been teaching for a number of years in the third grade and many had pursued advanced coursework. However, none of the teachers had any significant educational background in science, and while the teachers perceived themselves as employing a variety of teaching techniques in both their regular and science instruction, observation revealed a traditional teacher-centered format using questions from the low end of Bloom's taxonomy (1956).

There are a number of implications from this study. Staff development directly influences instructional practices in most cases. These instructional practices of teachers do, in turn, have a statistically significant and measurable impact on the performance of students.

According to Piaget (1972), third graders are concrete thinkers. They do not normally possess the ability to think abstractly. The ability of the third graders to perform proficiently or advanced-proficiently on the rubric-assessed, open-ended question requires the use of abstract thinking skills in the upper three levels of Bloom's taxonomy (1956). The fact that a significantly higher number of students in the experimental classes performed at these levels is directly related to the instruction of the experimental group of teachers, since third graders would arguably not have been able to build the abstract connections on their own to answer the open-ended question at the levels they did. The abstract level of thinking ability necessary to make these connections does not fully emerge until closer to the eighth grade. This type of thinking would have to be modeled for the third graders. The teacher must occupy the role of "metacognitive coach" and explicitly model and guide the third grade student through the thinking process needed to achieve such abstract outcomes. It is therefore important that quality staff development in the use of higher-order questioning strategies such as the kind demonstrated in this study be provided for teachers. Without such specific interventions, teachers are not or less able to guide elementary students toward the type of abstract thinking required to achieve the performance necessary to keep pace with the increasing demands for measurable student outcomes, and, as a direct consequence, students will fall behind (Guskey, 1999, 2000; Sparks, 1997).

This study has a bearing on the required New Jersey State assessments. The Elementary School Proficiency Assessment (ESPA), the Grade Eight Proficiency Assessment (GEPA), and the High School Proficiency Assessment (HSPA) all have open-ended questions throughout each of the sub-tests included in all three assessments. As demonstrated by this study, rote instructional strategies will not provide students with the skills necessary to answer the in-depth nature of these higher-order questions (Firestone, Camilli, Yerecko, Monfils, & Mayrowetz, 2000).

Teachers must instruct with the same level of higher-order methodology (metacognitive coaching) to provide students with ongoing practice for this type of assessment. There is a link between teacher instruction in the use of higher-order questions and methodology and the ability of students to perform in the proficient and advanced proficient categories on the state assessments. Students who are not instructed in this style, but rather with a rote, teacher-centered, traditional methodology, perform predominantly in the non- and partially proficient categories on an open-ended assessment (Darling-Hammond, 2000).

A further implication of this study is the need for direct classroom intervention by a knowledgeable individual to help guide staff with staff development interventions and to insure these new practices are successfully and routinely being utilized in their classroom practice. Teachers may or may not implement strategies they have learned at staff development sessions. The only way to verify that the desired instructional practices are actively in use in the classroom is through regular classroom visitation. Without some documentation of the process, teachers will tend not to change their practice readily. This is verifiable by the control teachers who received significant staff development in their district but have observable difficulty implementing it in their classrooms. This becomes a mandate in light of the outcome of this study where the direct beneficiaries of the implementation or lack of implementation of improved instructional practice in the classroom are the students (Darling-Hammond et al., 1983).

Recommendations for Future Research

Replication of this study, including an impartial observer conducting the research and/or using standardized assessment instruments, would strengthen it.

A more refined internal methodology, such as conducting all teacher training as well as all assessment of student responses in the classroom at the same time of day and including gender, age, and different grade levels, would add precise dimensions to a replication of this study. Inclusion of larger numbers of students followed longitudinally for a period of years would yield strong conclusions for classroom teaching. Although this study yielded no connection between I.Q., academic, or socio-economic background and performance of pupils when the use of higher order questions was incorporated into the instructional practice of teachers, each of these factors, separately or collectively, could be looked at in connection with staff development and the instructional methods of teachers.

References

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals* (Handbook I). New York: Longmans, Green.

Cardellichio, T., & Field, W. (1997). Seven strategies that encourage neural branching. *Educational Leadership, 54*(6), 33–36.

Caulfield-Sloan, M. (2001) *The effect of staff development of teachers in the use of higher order questioning strategies on third grade students' rubric science assessment performance.* Unpublished doctoral dissertation, Seton Hall University, New Jersey.

CTB/McGraw-Hill. (1997). *TerraNova Basic Multiple Assessments Plus.* Monterey, CA: Author.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1), 28–29.

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research, 53*, 285–237.

Doyle, W. (1985). Recent research on classroom management: Implications for teacher preparation. *Journal of Teacher Education, 36*(3), 31–35.

Firestone, W. A., Camilli, G., Yurecko, M., Monfils, L., & Mayrowetz, D. (2000). State standards, socio-fiscal context and opportunity to learn in New Jersey. *Education Policy Analysis Archives, 8*(35), 1.

Gagne, R. (1965). *Elementary science: A new scheme of instruction.* Syracuse, NY: Syracuse University Department of Science Teaching.

Good, T. L., & Brophy, J. E. (1986). *Educational psychology* (3rd ed.). New York: Longman.

Guskey, T. R. (1999). Moving from means to ends. *Journal of Staff Development, 20*(2), 48.

Guskey, T. R. (2000). *Evaluating professional development.* Thousand Oaks, CA: Corwin Press.

Piaget, J. (1972). *The child and reality.* Middlesex, England: Penguin Books.

Riverside. (1993). *Cognitive Abilities Test, form 5, level 2.* Itasca, IL: Author.

Rosenshine, B., & Furst, N. F. (1973). *The use of direct observation to study teaching* (2nd ed.). Chicago: Rand McNally.

Sparks, D. (1997, September). A new vision for staff development: Deep and continuous change is needed if teachers are to succeed in an era of high standards. *Principal*, 20–22.

Williamson, R. D. (1998). Designing diverse learning styles. *Schools in the Middle, 7*(4), 28–31.

Maryrose B. Caulfield-Sloan is the Principal at Jefferson School, Maplewood, New Jersey, and an Adjunct Professor at Seton Hall University, South Orange, New Jersey.

Mary F. Ruzicka is a Professor at Seton Hall University, South Orange, New Jersey.